

LSE Research Online

Colin Howson

Repelling a Prussian charge with a solution to a paradox of Dubins

**Article (Published version)
(Refereed)**

Original citation:

Howson, Colin (2016) *Repelling a Prussian charge with a solution to a paradox of Dubins*. [Synthese](#). ISSN 0039-7857

DOI: [10.1007/s11229-016-1205-y](https://doi.org/10.1007/s11229-016-1205-y)

Reuse of this item is permitted through licensing under the Creative Commons:

© 2016 The Author
CC BY 4.0

This version available at: <http://eprints.lse.ac.uk/67608/>

Available in LSE Research Online: September 2016

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

Repelling a Prussian charge with a solution to a paradox of Dubins

Colin Howson¹

Received: 2 May 2016 / Accepted: 20 August 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Pruss (Thought 1:81–89, 2012) uses an example of Lester Dubins to argue against the claim that appealing to hyperreal-valued probabilities saves probabilistic regularity from the objection that in continuum outcome-spaces and with standard probability functions all save countably many possibilities must be assigned probability 0. Dubins's example seems to show that merely finitely additive standard probability functions allow reasoning to a foregone conclusion, and Pruss argues that hyperreal-valued probability functions are vulnerable to the same charge. However, Pruss's argument relies on the rule of conditionalisation, but I show that in examples like Dubins's involving nonconglomerable probabilities, conditionalisation is self-defeating.

Keywords Probability · Hyperreal · Nonconglomerability · Conditionalisation

1 Introduction

Halfway through the last century Abraham Robinson showed that there is an elementary extension of the first-order structure of the ordered field of real numbers, containing infinitesimal¹ and, because of the field axioms, infinitely large numbers. The members of such a *nonstandard extension*, of which there are infinitely many

¹ Infinitesimals are the numbers in such an extension with absolute value smaller than every positive real number.

✉ Colin Howson
howson.colin@gmail.com

¹ London School of Economics and Political Science, London, UK

and infinitely many of different cardinalities,² are known as *hyperreals*. Because the extension is elementarily equivalent to the reals, any first-order sentence true in one is also true in the other, a result baptised ‘the Transfer Principle’. If the hyperreals are constructed ‘concretely’ as an ultrapower of the reals, the Transfer Principle is an immediate corollary of Łoś’s Theorem.

The hyperreals, and more generally a version of the cumulative hierarchy based on them (a rank *superstructure* of so-called internal sets), have turned out to be extremely useful in mathematical applications. Philosophers have also sought to exploit them in various ways, one of which has been an attempt to salvage the doctrine of probabilistic *regularity*. This is the demand that probabilities should be strictly positive, i.e. only the impossible event should receive zero probability, which runs up against the mathematical fact that only countably many members of an uncountable disjoint family can be assigned positive probability. Once the range of a probability function consists of some hyperreal extension, however, the objection no longer has force because all those outcomes can be assigned a positive infinitesimal probability, even in principle the same infinitesimal probability.³

This way of defending regularity has been challenged by Williamson (2007) and with different arguments by Pruss (2012, 2013) and Alan Hájek (2013). Williamson’s paper considers an infinitely-tossed fair coin to attempt a *reductio* from assuming that the probability of any infinite sequences of head-outcomes can be assigned an infinitesimal positive probability. The *reductio* proceeds by arguing that the sequence of outcomes starting from the second toss is isomorphic to that starting from the first and therefore both merit the same probability, from which a contradiction easily follows. Pruss (2013) contains an elegant mathematical argument which in effect extends to non-standard probability functions the fact mentioned earlier that for any standard probability function which includes in its domain a partition of the power of the continuum, only countably many cells can in principle have positive measure. Pruss shows that however (infinitely) big the hyperreal field H one selects to be one’s set of probability-values, there is always some algebra of events which for that choice of H will require infinitely many events to be assigned probability 0. Hájek shows that if the event-algebra is the power set of the nonstandard unit interval (in any given hyperreal extension), and any point receives a positive infinitesimal probability then the latter must exceed the length of some infinitesimal interval containing the point, violating, Hájek claims, possibly appropriate constraints (e.g. uniformity).

All these arguments have been challenged in the literature: in Williamson’s case by Weintraub (2008), who argues that it begs the question that isomorphism merits identical probabilities, and by Howson (2016) who argues that the claim of isomorphism is anyway mistaken; and in Pruss’s and Hájek’s cases by Hofweber (2014a, b). In both the Pruss and Hájek examples the choice of a sufficiently larger set of hyperreals could in principle restore regularity (in Hájek’s by admitting smaller infinitesimals

² The hyperreals are to this extent not as ‘real’ as the real numbers themselves, which are unique up to isomorphism in the standard model of set theory. The hyperreals can be seen as vindicating Leibniz’s view of infinitesimals as ideal elements facilitating standard calculations.

³ Bernstein and Wattenberg (1969), p. 176.

smaller than any in the original field),⁴ and in a careful discussion Hofweber argues that such a strategy is entirely compatible with a view of measurement in which the choice of value-range is not something that should be seen as fixed independently of the purpose it is intended to serve in any given case (2014b). Hájek responded (2013) by arguing that such an approach amounts to abandoning anything resembling a Kolmogorovian framework for probability which, he claims, requires that the range of the probability function must be fixed, and at the real numbers themselves since only for these is there an unambiguously meaningful notion of additivity (p. 20): countable additivity.⁵ This claim seems to me highly questionable. Ever since Abraham Robinson developed the modern theory of non-standard analysis, probabilists have been very successfully using hyperreal-valued probabilities, either defined on internal algebras, where hyperfinite additivity replaces countable, or on standard algebras, where the existence of a regular, hyperreal-valued finitely additive probability measure on an arbitrary Boolean algebra was proved first (to the best of my knowledge) sixty years ago by Nikodým.⁶ Moreover, for any specified real-valued measure there is a regular hyperreal-valued measure agreeing with it up to an infinitesimal.

These are all recognisably generalisations of the classical measure-theoretic framework in which Kolmogorov embedded probability theory⁷, with a stable notion of additivity for all hyperreal extensions, finite additivity. To insist that Kolmogorovian-style probability theory *must* employ countable additivity, and hence be defined on the real, runs counter not only to the belief of people like de Finetti and Savage, for whom only finite additivity is the correct form of additivity for Bayesian probability, but also possibly to that of Kolmogorov himself, who separated off countable additivity (in the form of an axiom of continuity) from the other, basic axioms, claiming no more for it than mathematical expediency.⁸ If I am correct, then, Hájek's parting challenge 'Just try providing an axiomatisation along the lines of Kolmogorov's that has any flexibility in the range built into it' (p. 22) has already been answered.⁹

It is also worth pointing out that the idea that Kolmogorov's axioms are to be regarded as written in stone, so to speak, is not one to which probabilists themselves

⁴ Though it is not obvious to me why a symmetry constraint thought to authorise uniformity over the reals should also be obliged to carry over to *infinitesimal* intervals.

⁵ Hyperreal probabilities are not countably additive because the usual limit procedure for defining infinite sums is not available (the least upper bound principle fails for hyperreal numbers).

⁶ Nikodým (1956). A simplified proof is in Luxemburg (1962).

⁷ By regarding a probability as a non-negative function on an algebra of events (which can always be identified with an algebra of subsets of a basic possibility-space, called by Kolmogorov a set of 'elementary events' (1950, p. 2).

⁸ 'We limit ourselves, arbitrarily, to only those models which satisfy Axiom VI (equivalent to countable additivity)' (1950, p. 15; emphasis in original). To inject a personal anecdote into the discussion, I regularly see in the philosophical literature references to 'Kolmogorov's axioms' in which only binary, and hence finite, additivity is assumed. One could see this as mere inaccuracy, but I think a more plausible view is that finite additivity is generally thought to be the fundamental additivity rule (it doesn't of course prohibit countable additivity).

⁹ Although I believe that these counters are successful, this should not be taken as a positive endorsement of attempts to save regularity by appeal to the hyperreals, and I will argue in the concluding section of this paper that trying to save regularity in this—or any other—way is misconceived.

by and large seem to subscribe, including, perhaps surprisingly, Kolmogorov himself. In a later paper (1995; published originally in 1948) he listed what he thought were the defects of the formalism presented in his original monograph (among them the presence of the underlying set of ‘elementary outcomes’ in which the algebra of events/subsets are defined, which he now regards as a purely ‘artificial superstructure’).¹⁰ In the decades following the publication of that monograph the discipline has seen many experimental variations on Kolmogorov’s original presentation: Kolmogorov’s own, just described; others in which probabilities are defined directly on formulas from an appropriate, possibly infinitary language (Scott and Krauss 1966); de Finetti’s approach (1974), which rejects countable additivity and in which coherent probabilities are defined on any arbitrary set of events (which must however include the certain event); non-standard probabilities; and some discussions of quantum mechanics in which *negative* probabilities have been seriously considered.

So much by way of introduction: I now turn to the main focus of this paper, the argument of Pruss (2012), exploiting a notorious problem with finitely additive standard probability functions: *non-conglomerability*. I will show that this argument, ingenious though it is, does not prove what its author claims it does.

2 Pruss’s argument

It is well known that merely finitely additive (standard) probability functions can always have as their domain the full power-set of a possibility-spaces, however big it might be. This is nice; not so nice is that in an infinite possibility-space S these probabilities are *non-conglomerable* along some countably infinite margin: that is to say, for any such function P there is a countably infinite partition $\{C_i : i \in \mathbb{Z}_+\}$ of S and real numbers a, b such that for some event D , $P(D|C_i) \in [a, b]$ for all $i \in \mathbb{Z}_+$ and $P(D) \notin [a, b]$; by contrast, countably additive probability functions are conglomerable in every countable margin¹¹. This feature of finitely additive functions was acknowledged by de Finetti, who first identified it but nevertheless still rejected the rule of countable additivity. What might seem even more problematic from his point of view is that any such probability function seems easily Dutch-Bookable and hence incoherent: one of the C_i must occur, suppose C_m , leaving the ‘owner’ of these probabilities in effect citing two different odds on D , namely $P(D|C_m)$ and $P(D)$. One of these is always larger or always smaller than the other. Under the usual assumption that this person is willing to bet indifferently on or against any proposition at their fair-betting rate (with the stakes in utiles), it is easy to see that they can be made to lose a positive amount in any case. But despite appearances there is no incoherence here. De Finetti, who introduced the concept of incoherence into probabilistic discourse and whom we should presumably therefore trust for a correct understanding of it, defined it as the

¹⁰ Its original function was presumably to support a class of random variables, but these can be identified with homomorphisms from the event algebra into the algebra of Borel sets.

¹¹ More precisely, a full conditional probability P on the power set of S is conglomerable along every countable partition of S just in case the unconditional function $P(\cdot | S)$ is countably additive.

vulnerability of a *finite* set of bets to a Dutch Book (1972, p. 79). In the present case, by contrast, infinitely many bets are required to ensure a certain loss.¹²

Incoherence may not be an issue for non-conglomerable probabilities, but in a letter to de Finetti Lester Dubins presented an example in which they appear to have genuinely pathological, if not paradoxical, consequences. Dubins's example consists of two procedures, A and B, which generate positive integers as the values of some random variable X, and A and B are selected each with the probability 1/2. It is assumed that the chance of any given integer n being selected by A is identically 0 (so the probability function is assumed to be only finitely additive), and the chance of it being selected by B is 2^{-n} .¹³ A simple Bayes's Theorem calculation shows that $P(A|X = n) = 0$ and $P(B|X = n) = 1$ for every $n \in \mathbb{Z}_+$ (read A and B as corresponding propositions saying that A is selected, B is selected). Thus we have the nonconglomerability of P along the margin $\{X = n : n = 1, 2, 3, \dots\}$ since $P(A) = P(B) = 1/2$. Assuming that you were to adopt the Bayesian rule of conditionalisation after observing X, you would be certain that it is B that was selected *whichever integer had been observed*. The paradoxical character of the example lies in the fact that, as has just been done, you can work all this out in advance of observing X; in Kadane et al.'s nice phrase (1996), you can 'reason to a foregone conclusion'. This is clearly not a formal paradox, i.e. a contradiction (modulo the rules of finitely additive probability). But it is a paradox in the informal sense that it challenges a basic intuition, namely that you can predict with certainty exactly *how* your belief will change, and in this case change radically, before you are even aware of the evidence that will cause the change.

Though such fatalistic reasoning is, in the context of standard probability functions and the rule of conditionalisation, a consequence of rejecting countable additivity, it is, as Pruss points out, easily reproducible in a nonstandard setting. Wenmackers and Horsten (2013) have presented one method (based on what they call a 'numerosity' measure) of assigning infinitesimal probabilities to the positive integers in which they (hyperfinitely) sum to 1. So suppose that procedure A above assigns such an infinitesimal probability to the (standard) integer n, with B assigning the value 2^{-n} as before. Involving as they do only arithmetical operations, similar Bayes's Theorem calculations to those above show that $P(A|X = n)$ is infinitesimal for every n, while $P(B|X = n)$ is infinitely close to 1 for all n. Again, you know in advance that conditionalisation will give you these updated probabilities whichever n is generated: you will be practically certain that n was generated by B. Pruss's conclusion is that if you look to infinitesimals to escape failures of regularity you will end up with something as, or even more, unpalatable.

¹² The apparently ad hoc finiteness condition does have an important proof-theoretic rationale: another of de Finetti's major results (he called it 'the fundamental theorem') is that any coherent probability on any arbitrary set of events containing the certain event can be extended to any including algebra, including a power-set algebra (1974, p. 112). This is of course in sharp contrast to countably additive probability measures, where for example not all subsets of $[0,1]$ are measurable.

¹³ In his discussion of the example (1972, p. 205, 206), de Finetti suggested the procedure of flipping a fair coin until the first tail comes up as a model for B, and a fair infinite lottery as a model for A.

3 The diagnosis

Given either the assumption of merely finite additivity for standard probabilities, or the extension of the range of standard probability functions to hyperreal values, conditionalisation is the necessary and sufficient condition for the generation of these apparently paradoxical results. In other words if and only if you know that you will conditionalise, given those other assumptions, can you say that you know in advance how your belief *will* change. Thus neither finite additivity alone nor the assumption of hyperreal probability-values alone is culpable: the pathology only emerges with the complicity of conditionalisation. However, a little reflection arguably shows that given those same assumptions, conditionalisation is actually an *inappropriate* updating strategy: if you know that whichever value of X might be observed, conditionalisation on it will lead you to conclude that A is practically certain, *and you nevertheless maintain your prior of $1/2$* , then it would seem that you have implicitly rejected conditionalisation as an appropriate policy (since the implicit possibility-space comprises both ‘possible worlds’ determining which of A and B are selected and also the possible values of X , that prior already implicitly considers the later). Jos Uffink observed in a different context that ‘if it is certain beforehand that a probability value will be revised downward, this value must have been too high to start with, and could not have been a faithful representation of our opinion’ (1996, p. 68). Conversely, if the prior for A is a faithful representation of our opinion, and its revision downward is due to the presence of a particular assumption or rule of inference (the rule in the present case being conditionalisation), then the assumption is false or the rule invalid or otherwise inappropriate. This is not at all to impugn the inference via Bayes’s Theorem to the conditional probabilities $P(A|X = n) = 0$, $P(B|X = n) = 1$, which of course remains valid since it is a condition of coherence. What is impugned is the appropriateness of conditionalising on those probabilities (as we shall see shortly, in his own discussion of this example de Finetti himself came close to explicitly rejecting conditionalisation).

Here, then, we have a situation where finite additivity or hyperreal probabilities appear to generate a paradoxical conclusion, but only in conjunction with an updating rule, conditionalisation, which in this context seems to be self-undermining.¹⁴ According to a well-known argument, however, adopting any other policy falls prey to incoherence: anyone announcing a different updating strategy, and is willing to bet either way at their fair-betting rate, is easily shown to be vulnerable to what is called in the Bayesian vernacular a ‘dynamic’ Dutch Book. In the present context we can forget the qualifier ‘dynamic’, since the Dutch Book to which a non-conditionalising strategy in the present context is vulnerable is a quite ordinary, ‘synchronic’ one. Suppose my updated degree of belief in A after observing that $X=n$ is some number q_n , where q_n is positive for every n , while, of course, $P(A|X = n) = 0$. Then given that only one value of X will be recorded, anyone who is willing to bet indifferently on or against A at any or all their fair-betting quotients can of course be Dutch Booked, using the same sort of Dutch Book as in Sect. 2. Unfortunately (or not) as with that

¹⁴ Kadane et al. state that repudiating conditionalisation is just one possible way, among others, to avoid that conclusion (1996, p. 1235), but if the foregoing is correct the prior distribution over A and B implicitly rules it out anyway.

earlier Dutch Book this one also is no evidence of incoherence, since as with the first it requires an infinity of bets being made, and as we know from the earlier discussion the incoherence of any set of betting quotients requires that a *finite* subset must be Dutch Bookable.

This is of course not the only argument in the literature for conditionalisation. Another popular candidate appeals to expectations: there is a well-known theorem that any act which maximises utility relative to a conditionalisation-updated belief function has at least as great a utility with respect to your prior distribution as a utility-maximising act computed relative to any other updating rule, and is strictly greater if there is a unique such act. In other words, conditionalisation maintains, in this sense, your current utility ranking. A corresponding ‘non-pragmatic’ result transfers the expectations in question from utilities to measures of inaccuracy,¹⁵ subject to an important condition (the belief function must be ‘immodest’).^{16,17} It might of course be questioned why it should be automatically desirable to have your current expectations/estimates preserved in this way on learning new information. Be that as it may, there is a more immediate objection: these results, pragmatic and non-pragmatic, are proved for the restricted case where the possible evidence reports to be updated on form a finite partition. Easwaran has shown (2013) that they are preserved in the infinite case if the probability function is countably additive. In the present context, of course, the information updated on is a member of the countable partition $\{X = i : i = 1, 2, \dots\}$ and the probability function is not countably additive.¹⁸

And in the context of Dubins’s example the (allegedly) optimising character of conditionalisation-updated belief functions does indeed fail to be preserved: de Finetti himself in effect presented an infinite family of posterior distributions yielding superior expected gains by comparison with that generated by conditionalising, when judged from the prior standpoint. In his discussion of the Dubins example (1972, p. 205) he pointed out that the strategy of betting on B if the observed value n is less than or equal to k , and on A otherwise, gives a probability of winning, namely $1 - 2^{-k}$, that tends to 1 as k tends to infinity (it is not difficult to see that varying the value of the prior probability within $(0,1)$ doesn’t change the limit probability of 1). There is no optimal member of this family, and setting the value of k at infinity (bet on all finite values) is just the strategy based on conditionalising (always bet on B whatever the value of X observed) which as judged from the prior standpoint only a half chance of winning, whereas for all values of k greater than 1 that chance exceeds $1/2$. Judged from the prior standpoint, the betting strategy recommended by conditionalisation thus has a smaller expected value than the other. Implicit in the latter is a family of

¹⁵ Of the type James Joyce introduced into the Bayesian literature; see Joyce (2009).

¹⁶ Greaves and Wallace (2006).

¹⁷ A function is immodest if it assigns maximum expected accuracy just to itself. The well-known Brier score is one such immodest measure.

¹⁸ There are other purported justifications of conditionalisation, but they are to my mind all question-begging in one way or another. For example, another often-canvassed candidate appeals to the fact that minimising Kullback–Leibler divergence, subject to the constraint that the new evidence E is learned with probability 1, selects $P(\cdot|E)$ as the closest probability function to the prior $P(\cdot)$. K–L divergence is a (directed) distance measure in function space, but as such it is far from unique, and in any case it is not clear why minimising distance from the prior distribution is the appropriate criterion of choice for the posterior.

belief functions $P_{k,n}(B)$, constant on the set $\{n : n < k\}$ and dropping to some small value for large k . In fact, such a function is uniquely determined by the condition that the prior expectation of any bet at the corresponding odds is zero: then we must have $P_n(B) = P(B|X \leq k) = 1$ if $n \leq k$ and equals $P(B|X > k) = (1 + 2^k)^{-1}$ if $n > k$.¹⁹ This is of course not the only such updating function for B : another which is nowhere constant and satisfies the same intuition that smaller values count more heavily in favour of B is the function $P_n(B) := P(X = n|B) = 2^{-n}$.

All in all, then, it seems that the verdict must be that the Dubins example fails to demonstrate Pruss's claim that although infinitesimal probabilities are mathematically well-defined they

allow one to have an event whose credence ...is significantly less than one half and a set-up where no matter what results, one will end up within an infinitesimal of certainty that the event occurred (2012, p. 9)

A claim which anyway is strictly speaking incorrect. Infinitesimal probabilities do not generate that pathological result by themselves, any more than a standard non-conglomerable probability function does: they generate it only in conjunction with conditionalisation. But since conditionalisation is implicitly denied where non-conglomerable probabilities are involved, any result obtained by its means can be viewed as akin to one produced trivially from an inconsistency. Hyperreal-valued probability functions behave in many ways like finitely additive standard functions, and it has often been observed that the class of merely finitely additive probabilities is a sort of wild place where the comfortable certainties of countably additive functions no longer hold. But I think that up to now it has not been generally realised that, as far as epistemic probability is concerned, conditionalisation is one of them.²⁰

4 Conclusion

If the foregoing is correct then Pruss's objection may fail to show that appealing to hyperreal probabilities to salvage regularity brings in its train reasoning to a foregone conclusion, but it certainly has other consequences scarcely more palatable. As to regularity, it is far from clear to me what dividends if any accrue to salvaging it by appeal to the hyperreals, other than the mathematical satisfaction of knowing that however many possible contingent events we can conjure, there will always be positive numbers enough to measure them. By comparison with what has to be foregone that seems scant justification, a sort of *l'art pour l'art* when compared with the wealth of results of classical mathematical probability on the reals in which regularity is

¹⁹ This function, or family of functions, is noted by Kadane et al. (1996, p. 1232), who also point out that taking k as infinite leads back to the conditionalised updating function, though they nowhere mention de Finetti's own discussion.

²⁰ This doesn't imply that conditionalisation is problematic in typical Bayesian inductive scenarios, whether in the context of countable additivity, where of course its use seems entirely straightforward, or even in the finitely additive context where examples like Dubins's are highly atypical and can be regarded as isolated pathologies. Even under finite additivity quite strong versions of the classical limit theorems remain provable (see for example Chen 1977).

violated systematically, not just in the powerful ‘almost surely’ theorems but also in the ubiquitous continuous distributions which make up the bulk of applied probability.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bernstein, A. R., & Wattenberg, F. (1969). Nonstandard measure theory. In W. A. J. Luxemburg (Ed.), *Applications of model theory to algebra, analysis and probability* (pp. 171–186). New York: Holt, Rinehart and Winston.
- Chen, R. (1977). On almost sure convergence in a finitely additive setting. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 37, 341–356.
- de Finetti, B. (1972). *Probability, induction and statistics*. New York: Wiley.
- de Finetti, B. (1974). *Theory of probability* (Vol. 1). New York: Wiley.
- Easwaran, K. (2013). Expected accuracy supports conditionalization and conglomerability and reflection. *Philosophy of Science*, 80, 119–142.
- Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115, 607–632.
- Hájek, A. (2013). Staying regular, unpublished manuscript.
- Hofweber, T. (2014a). Infinitesimal chances. *Philosophers' Imprint*, 14, 1–34.
- Hofweber, T. (2014b). Cardinality arguments against regular probability measures. *Thought*, 3, 166–175.
- Howson, C. (2016). Regularity and infinitely tossed coins. *European Journal for the Philosophy of Science* (online first).
- Joyce, J. M. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief: Synthese library* (pp. 263–297). New York: Springer.
- Kadane, J. B., Schervish, M. J., & Seidenfeld, T. (1996). Reasoning to a foregone conclusion. *Journal of the American Statistical Association*, 91, 1228–1235.
- Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. New York: Chelsea.
- Kolmogorov, A. N. (1995). Complete metric boolean algebras. *Philosophical Studies* 77, 57–66 (R. Jeffrey of ‘Algèbres de Boole métriques complètes’ (1948), Trans.).
- Luxemburg, W. A. J. (1962). Two applications of the method of construction by ultrapowers to analysis. *Bulletin of the American Mathematical Society*, 68, 416–419.
- Nikodým, O. (1956). On extension of a given finitely-additive field-valued, non-negative measure, on a finitely additive Boolean tribe, to another tribe more ample. *Rendiconti del Seminario Matematico della Università di Padova*, 26, 232–327.
- Pruss, A. R. (2012). Infinite lotteries, perfectly thin darts and infinitesimals. *Thought*, 1, 81–89.
- Pruss, A. R. (2013). Probability, regularity, and cardinality. *Philosophy of Science*, 80, 231–240.
- Scott, D., & Krauss, P. (1966). Assigning probabilities to logical formulas. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 219–264). Amsterdam: North Holland.
- Uffink, J. (1996). The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of Modern Physics*, 27B, 47–81.
- Wenmackers, S., & Horsten, L. (2013). Fair infinite lotteries. *Synthese*, 190, 37–61.
- Weintraub, R. (2008). How probable is an infinite sequence of heads? *Analysis*, 68, 247–250.
- Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, 67, 173–180.